Quantum Gibbs Sampling Using Szegedy Operators

Robert R. Tucci
P.O. Box 226
Bedford, MA 01730
tucci@ar-tiste.com

January 14, 2010

Abstract

We present an algorithm for doing Gibbs sampling on a quantum computer. The algorithm combines phase estimation for a Szegedy operator, and Grover's algorithm. For any $\epsilon > 0$, the algorithm will sample a probability distribution in $\mathcal{O}(\frac{1}{\sqrt{\delta}})$ steps with precision $\mathcal{O}(\epsilon)$. Here δ is the distance between the two largest eigenvalue magnitudes of the transition matrix of the Gibbs Markov chain used in the algorithm. It takes $\mathcal{O}(\frac{1}{\delta})$ steps to achieve the same precision if one does Gibbs sampling on a classical computer.

1 Introduction

In Ref.[1], Szegedy proposed a quantum walk operator for each classical Markov chain. In Ref.[2], Somma et al. proposed a method for doing simulated annealing on a quantum computer. In Ref.[3], Wocjan et al. showed how to improve the Somma et al. algorithm. The algorithms of Somma et al. and Wocjan et al. both use Szegedy operators. In Ref.[4], I presented computer programs called QuSAnn and Multiplexor Expander that implement ideas of Refs.[2] and [3], and also some of my own ideas about quantum multiplexors.

In Ref.[5], I described one particular algorithm for doing Gibbs and Metropolis-Hastings sampling of a classical Bayesian network (i.e., a probability distribution) on a quantum computer. In this paper, I describe a different algorithm for doing Gibbs sampling on a quantum computer. Unlike my first algorithm, this one uses Szegedy operators. For any $\epsilon > 0$, this new algorithm will sample a Bayesian network in $\mathcal{O}(\frac{1}{\sqrt{\delta}})$ steps with precision $\mathcal{O}(\epsilon)$. Here δ is the distance between the two largest eigenvalue magnitudes of the transition matrix of the Gibbs Markov chain used in the algorithm. It takes $\mathcal{O}(\frac{1}{\delta})$ steps to achieve the same precision if one does Gibbs sampling on a classical computer.

This paper assumes that its reader has read the section entitled "Notation and Preliminaries" in Ref.[5]. The reader should refer to Refs.[5, 4] for clarification when any notation of this paper eludes him.

2 Dual Gibbs Markov Chains

In this section, we will discuss dual "Gibbs" Markov chains with transition matrices M_1 and M_2, respectively. These two transition matrices are both defined in terms of a single classical Bayesian network $\underline{x}$.

2.1 Definitions of M_1 and M_2

Consider a classical Bayesian net with N_{nds} nodes, labeled $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{N_{nds}}$ where $\underline{x}_j \in S_{\underline{x}_j}$ for each j. (As usual in my papers, I indicate random variables by underlining them.) Let $\underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{N_{nds}})$. Let $\underline{x}$ assume values in a set $S_{\underline{x}}$ which has $N_S = 2^{N_B}$ elements.

Let

$$\pi(x) = P(\underline{x} = x) \tag{1}$$

for all $x \in S_{\underline{x}}$.

For $N_{nds} = 3$ and $x, y \in S_{\underline{x}}$, let

$$M_1(y|x) = P_{\underline{x}_1|\underline{x}_2\underline{x}_3}(y_1|x_2, x_3) P_{\underline{x}_2|\underline{x}_3\underline{x}_1}(y_2|x_3, y_1) P_{\underline{x}_3|\underline{x}_1\underline{x}_2}(y_3|y_1, y_2) \ , \tag{2}$$

and

$$M_2(y|x) = P_{\underline{x}_1|\underline{x}_2\underline{x}_3}(y_1|y_2,y_3)P_{\underline{x}_2|\underline{x}_3\underline{x}_1}(y_2|y_3,x_1)P_{\underline{x}_3|\underline{x}_1\underline{x}_2}(y_3|x_1,x_2)\ . \tag{3}$$

($M_2(y|x)$ can be obtained by swapping x_i and y_i in the *conditioned* arguments of $M_1(y|x)$.) Note that $\sum_y M_1(y|x) = 1$ and $\sum_y M_2(y|x) = 1$. Define M_1 and M_2 for arbitrary N_{nds} using the same pattern. M_1 and M_2 are transition matrices of the type typical for Gibbs sampling. (See Ref.[5] for an introduction to Gibbs sampling and the more general Metropolis-Hastings sampling).

You can check that $\pi()$ is not a detailed balance of either M_1 nor M_2 separately. However, the following property is true. We will refer to this property by saying that $\pi()$ is a detailed balance of the pair (M_1, M_2).

Claim 1

$$M_1(y|x)\pi(x) = M_2(x|y)\pi(y) \tag{4}$$

for all $x, y \in S_{\underline{x}}$.

proof:

Let $P(x_j, x_k, \ldots)$ stand for $P(\underline{x}_j = x_j, \underline{x}_k = x_k, \ldots)$. Assume $N_{nds} = 3$ to begin with. One has

$$\begin{aligned}
\frac{M_1(y|x)}{M_2(x|y)} &= \frac{P(y_1|x_2,x_3)P(y_2|x_3,y_1)P(y_3|y_1,y_2)}{P(x_1|x_2,x_3)P(x_2|x_3,y_1)P(x_3|y_1,y_2)} &(5)\\
&= \frac{P(y_1,x_2,x_3)P(y_2,x_3,y_1)P(y_3,y_1,y_2)}{P(x_1,x_2,x_3)P(x_2,x_3,y_1)P(x_3,y_1,y_2)} &(6)\\
&= \frac{P(y_1,x_2,x_3)P(y_1,y_2,x_3)P(y)}{P(x)P(y_1,x_2,x_3)P(y_1,y_2,x_3)} &(7)\\
&= \frac{P(y)}{P(x)}\ . &(8)
\end{aligned}$$

A proof for an arbitrary number N_{nds} of nodes follows the same pattern.
QED

2.2 Eigenvalues of M_1, M_2 and M_{hyb}

Let

$$\Lambda_j(y|x) = \sqrt{M_j(y|x)}\ , \tag{9}$$

for $j = 1, 2$ and $x, y \in S_{\underline{x}}$. It's convenient to define a hybrid function of M_1 and M_2, as follows:

$$M_{hyb}(y|x) = \Lambda_2(x|y)\Lambda_1(y|x) \tag{10}$$

for $x, y \in S_{\underline{x}}$. (Note that unlike $M_1(y|x)$ and $M_2(y|x)$, $M_{hyb}(y|x)$ is not a probability function in y, its first argument.)

Define the quantum states

$$|(\pi)^\eta\rangle = \sum_x [\pi(x)]^\eta |x\rangle \tag{11}$$

for $\eta = \frac{1}{2}, 1$. (Note that only the $\eta = \frac{1}{2}$ state is normalized in the sense of quantum mechanics.)

Claim 2

$$M_j|\pi\rangle = |\pi\rangle \;\; \textit{for } j = 1, 2 \;, \tag{12}$$

and

$$M_{hyb}|\sqrt{\pi}\rangle = |\sqrt{\pi}\rangle \;. \tag{13}$$

Also, M_1, M_2 and M_{hyb} have the same eigenvalues.

proof:

Taking the square root of both sides of the pair detailed balance statement Eq.(4), we get

$$\Lambda_1(y|x)\sqrt{\pi(x)} = \Lambda_2(x|y)\sqrt{\pi(y)} \;. \tag{14}$$

Therefore,

$$M_{hyb}(y|x) = \Lambda_2(x|y)\frac{1}{\sqrt{\pi(x)}}\Lambda_2(x|y)\sqrt{\pi(y)} \;\; = \;\; \frac{1}{\sqrt{\pi(x)}}M_2(x|y)\sqrt{\pi(y)} \;. \tag{15}$$

Hence,

$$\sum_x M_1(y|x)\pi(x) = \sum_x M_2(x|y)\pi(y) = \pi(y) \;, \tag{16}$$

$$\sum_y M_2(x|y)\pi(y) = \sum_y M_1(y|x)\pi(x) = \pi(x) \;, \tag{17}$$

and

$$\sum_x M_{hyb}(y|x)\sqrt{\pi(x)} = \sum_x \frac{1}{\sqrt{\pi(x)}}M_2(x|y)\sqrt{\pi(y)}\sqrt{\pi(x)} = \sqrt{\pi(y)} \;. \tag{18}$$

Order the elements of the finite set $S_{\underline{x}}$ in some preferred way. Use this preferred order to represent M_1, M_2 and M_{hyb} as matrices. Define a diagonal matrix D whose

diagonal entries are the numbers $\pi(x)$ for each $x \in S_{\underline{x}}$, with the x ordered in the preferred order:

$$D = diag[(\pi(x))_{\forall x}] \ . \tag{19}$$

Since

$$M_2^T = D^{-1} M_1 D \ \ , M_{hyb}^T = D^{-\frac{1}{2}} M_2 D^{\frac{1}{2}} \ , \tag{20}$$

it follows that

$$\det(M_1 - \lambda) = \det(M_2 - \lambda) = \det(M_{hyb} - \lambda) \tag{21}$$

for any $\lambda \in \mathbb{C}$.
QED

Let the eigenvalues[1] of M_{hyb} (and also of M_1 and M_2) be $m_0, m_1, \ldots m_{N_S-1} \in \mathbb{C}$ with $m_0 = 1 > |m_1| \geq |m_2| \ldots \geq |m_{N_S-1}|$. Define $|m_j\rangle$ to be the corresponding eigenvectors of M_{hyb} (but not necessarily of M_1 and M_2). Thus

$$M_{hyb}|m_j\rangle = m_j|m_j\rangle \ , \tag{22}$$

for $j = 0, 1, \ldots, N_S - 1$. In particular, $|m_0\rangle = |\sqrt{\pi}\rangle$.

For each j, define $\varphi_j \in [0, \frac{\pi}{2}]$ and $\eta_j \in [0, 2\pi)$ so that $m_j = e^{i\eta_j} \cos\varphi_j$. (Thus, $\cos\varphi_j \geq 0$). Note that $m_0 = 1$ so $\varphi_0 = 0$. The M_1 eigenvalue gap δ is defined as $\delta = 1 - |m_1|$. $\delta \approx \frac{\varphi_1^2}{2}$ when φ_1 is small.

3 Q-Embeddings U_1 and U_2

In this section, we will define a "q-embedding" U_j of M_j, for $j = 1, 2$. (For more information about q-embeddings, see Ref.[5].)

For simplicity, we begin this section by considering a Bayesian net with only 3 nodes $\underline{x}_1, \underline{x}_2, \underline{x}_3$, and such that each of these nodes is binary (i.e., $S_{\underline{x}_j} = Bool$ for $j = 1, 2, 3$). At the end of this section, we will show how to remove these restrictions and make our treatment valid for general Bayesian networks.

Using the same language as Ref.[5], consider a unitary matrix U_1 of the form shown in Fig.1, with its multiplexor gates defined as follows. Let $\underline{x}_j\langle k\rangle \in Bool$ and $\underline{x}'_j\langle k\rangle \in Bool$ for any j, k. U_1 has 3 analogous gates (a.k.a. nodes) labeled $(\underline{x}'_1\langle 2\rangle, \underline{x}_3\langle 2\rangle, \underline{x}_2\langle 2\rangle)$, $(\underline{x}'_2\langle 3\rangle, \underline{x}'_1\langle 3\rangle, \underline{x}_3\langle 3\rangle)$, and $(\underline{x}'_3\langle 4\rangle, \underline{x}'_2\langle 4\rangle, \underline{x}'_1\langle 4\rangle)$. Consider the first of these for definiteness. Let the probability amplitude $A(x'_1\langle 2\rangle, x_3\langle 2\rangle, x_2\langle 2\rangle | x'_1\langle 1\rangle, x_3\langle 1\rangle, x_2\langle 1\rangle)$ of node $(\underline{x}'_1\langle 2\rangle, \underline{x}_3\langle 2\rangle, \underline{x}_2\langle 2\rangle)$ satisfy the constraint

[1] There must be a single eigenvalue 1 and all others must have a magnitude strictly smaller than one because of the Frobenius-Perron Theorem. The eigenvalues may be complex.

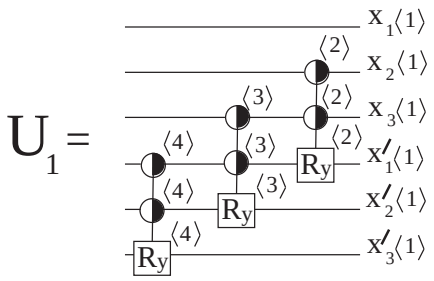


Figure 1: Unitary matrix U_1 expressed as a product of quantum multiplexors.

$$A(x_1'\langle 2\rangle, x_3\langle 2\rangle, x_2\langle 2\rangle | x_1'\langle 1\rangle = 0, x_3\langle 1\rangle, x_2\langle 1\rangle) \tag{23}$$

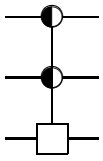
$$= \sqrt{P_{\underline{x}_1|\underline{x}_3,\underline{x}_2}(x_1'\langle 2\rangle | x_3\langle 2\rangle, x_2\langle 2\rangle)}\,\delta^{x_2\langle 1\rangle}_{x_2\langle 2\rangle}\delta^{x_3\langle 1\rangle}_{x_3\langle 2\rangle}\ . \tag{24}$$

If we indicate non-zero entries by a plus sign,

$$A \quad = \tag{25}$$

$(x_1', x_3, x_2)=$	000	001	010	011	⋯
000	+				⋯
001		+			⋯
010			+		⋯
011				+	⋯
100	+				⋯
101		+			⋯
110			+		⋯
111				+	⋯

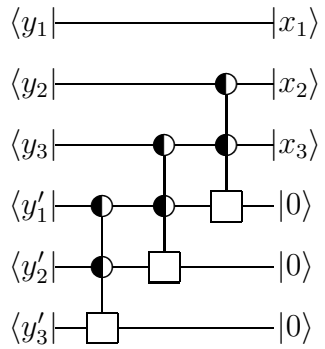
$$\rightarrow \sum_{\vec{b}\in Bool^2} e^{i\theta_{\vec{b}}\sigma_Y} \otimes P_{\vec{b}} = \tag{26}$$



for some $\theta_{\vec{b}} \in \mathbb{R}$. Here the right pointing arrow means that the expression at the origin of the arrow can be extended to the expression at the target of the arrow.

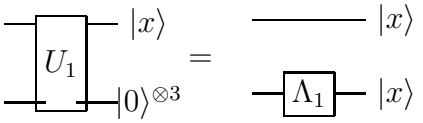
From the above definition of U_1, it follows that, for $x, x', y, y' \in Bool^3$,

$$\begin{matrix}\langle y| \\ \langle y'|\end{matrix} \; U_1 \; \begin{matrix}|x\rangle \\ |0\rangle^{\otimes 3}\end{matrix} \quad = \quad$$



$$= \quad \Lambda_1(y'|x)\delta(y,x) \; . \tag{27, 28}$$

Hence,



$$\text{or } U_1|0\rangle^{\otimes 3}|x\rangle = (\Lambda_1|x\rangle)|x\rangle \; . \tag{29}$$

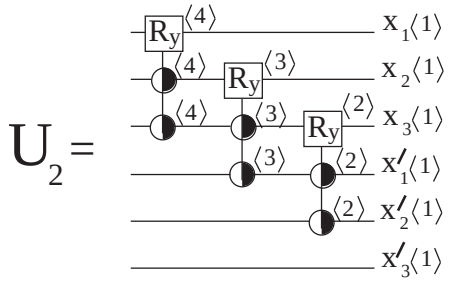
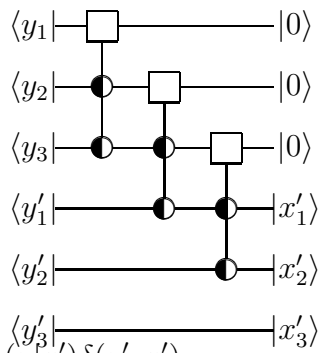


Figure 2: Unitary matrix U_2 expressed as a product of quantum multiplexors.

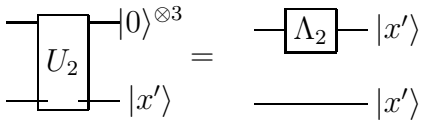
Besides U_1, it is convenient to consider another unitary matrix called U_2. We define U_2 to be of the form of Fig.2, where the multiplexors are defined in such a way that U_2 satisfies, for all $x, x', y, y' \in Bool^3$,

$$\begin{matrix}\langle y| \\ \langle y'|\end{matrix} \; U_2 \; \begin{matrix}|0\rangle^{\otimes 3} \\ |x'\rangle\end{matrix} \quad = \quad$$



$$= \quad \Lambda_2(y|x')\delta(y',x') \; . \tag{30, 31}$$

Hence



$$\text{or } U_2|x'\rangle|0\rangle^{\otimes 3} = |x'\rangle(\Lambda_2|x'\rangle)\ . \tag{32}$$

U_j is called the q-embedding of M_j for $j = 1, 2$.

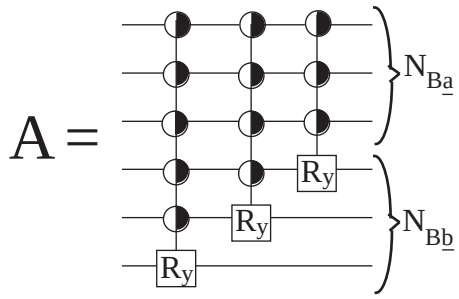


Figure 3: The unitary matrix A is a quantum embedding of a probability matrix $P(b|a)$, where a has $N_{B\underline{a}}$ bits and b has $N_{B\underline{b}}$ bits.

So far we have considered the q-embeddings U_1 and U_2 for the case of a classical Bayesian network $\underline{x}$ with 3 binary nodes. What if $\underline{x}$ has N_{nds} nodes and some of those nodes have more than 2 states? In that case, we must use several qubits (horizontal lines) for each node $\underline{x}_i$ (and an equal number of qubits for the dual node $\underline{x}'_i$) in Figs.1 and 2. More specifically, suppose $P(x_1|x_2, x_3, \ldots, x_{N_{nds}})$ equals $P(b|a)$ where $a \in Bool^{N_{B\underline{a}}}$ and $b \in Bool^{N_{B\underline{b}}}$. For the number of bits $N_{B\underline{a}}$, define the number of states $N_{S\underline{a}} = 2^{N_{B\underline{a}}}$. Likewise, let $N_{S\underline{b}} = 2^{N_{B\underline{b}}}$. The constraint Eq.(24) generalizes to

$$A(b, \tilde{a}|\tilde{b} = 0, a) = \sqrt{P(b|a)}\delta_a^{\tilde{a}}\ , \tag{33}$$

where $a, \tilde{a} \in Bool^{N_{B\underline{a}}}$ and $b, \tilde{b} \in Bool^{N_{B\underline{b}}}$. Eq.(33) can be expressed in matrix form as follows:

$$[A(b, \tilde{a}|\tilde{b} = 0, a)] = \begin{array}{c|c} & (\tilde{b} = 0, a) \rightarrow \\ \hline (b, \tilde{a}) & D^{0,0} \\ \downarrow & D^{1,0} \\ & \cdots \\ & D^{N_{S\underline{b}}-1,0} \end{array}\ , \tag{34}$$

where, for all $b \in Bool^{N_{B\underline{b}}}$, $D^{b,0} \in \mathbb{R}^{N_{S\underline{a}} X N_{S\underline{a}}}$ are diagonal matrices with entries

$$(D^{b,0})_{a,\tilde{a}} = \sqrt{P(b|a)}\delta_a^{\tilde{a}}\ . \tag{35}$$

By adding more columns to the matrix of Eq.(34), one can extended it (see section entitled "Q-Embeddings" in Ref.[5]) to a square matrix which can be expressed in terms of multiplexors as in Fig.3.

The Markov Blanket $MB(i)$ for a node $\underline{x}_i$ of the classical Bayesian network $\underline{x}$ satisfies (see section entitled "Notation and Preliminaries" in Ref.[5])

$$P(x_i|x_{\{i\}^c}) = P(x_i|x_{MB(i)}) \ . \tag{36}$$

If the set $MB(i)$ is strictly smaller than the set $\{i\}^c$, this property can be used to reduce the number of controls for the multiplexor in U_1 and U_2 corresponding to $P(x_i|x_{\{i\}^c})$.

Given the two q-embeddings U_1 and U_2 for a Bayesian network $\underline{x}$, we can define a unitary matrix U as follows

$$U = U_2^\dagger U_1 \ . \tag{37}$$

Matrix U has the following highly desirable property:

Claim 3 *For any $j, k \in \{0, 1, \ldots, N_S - 1\}$,*

$$\begin{matrix} \langle 0| \\ \langle m_j| \end{matrix} \ U \ \begin{matrix} |m_k\rangle \\ |0\rangle \end{matrix} \ = m_j \delta_j^k \ . \tag{38}$$

proof:

$$\begin{matrix} \langle 0| \\ \langle m_j| \end{matrix} \ U_2^\dagger U_1 \ \begin{matrix} |m_k\rangle \\ |0\rangle \end{matrix} = \sum_{y,x} \langle m_j|y\rangle \begin{bmatrix} \langle y|\Lambda_2^T \\ \langle y| \end{bmatrix} \begin{bmatrix} |x\rangle \\ \Lambda_1|x\rangle \end{bmatrix} \langle x|m_k\rangle \tag{39}$$

$$= \sum_{y,x} \langle m_j|y\rangle \Lambda_2^T(y|x)\Lambda_1(y|x)\langle x|m_k\rangle \tag{40}$$

$$= \langle m_j|M_{hyb}|m_k\rangle = m_j\delta_j^k \ . \tag{41}$$

QED

4 Szegedy Quantum Walk Operator W

In this section, we will define a special type of Szegedy quantum walk operator W corresponding to a Bayesian net $\underline{x}$. We will then find the eigenvalues of W.

4.1 Definition of W

As in Ref.[4], define the projection operator $\hat{\pi}$ and its dual projection operator $\check{\pi}$ by

$$\hat{\pi} = \begin{matrix} |0\rangle\langle 0| \\ \hline \end{matrix} \quad , \quad \check{\pi} = \updownarrow \hat{\pi} \updownarrow = \begin{matrix} \overline{} \\ |0\rangle\langle 0| \end{matrix} \quad . \tag{42}$$

Then the Szegedy quantum walk operator W for the Bayesian net $\underline{x}$ is defined by

$$W = U(-1)^{\check{\pi}} U^\dagger (-1)^{\hat{\pi}} \; . \tag{43}$$

4.2 Eigenvalues of W

To find the eigenvalues of W, we will use the following identities.

Claim 4

$$\hat{\pi}|m_j 0\rangle = |m_j 0\rangle \; , \tag{44a}$$

$$\hat{\pi}(U \updownarrow)|m_j 0\rangle = m_j |m_j 0\rangle \; , \tag{44b}$$

$$\hat{\pi}(\updownarrow U^\dagger)|m_j 0\rangle = m_j^* |m_j 0\rangle \; , \tag{44c}$$

for all $j \in \{0, 1, \ldots, N_S - 1\}$.

proof:

From the definition of $\hat{\pi}$, we see that

$$\hat{\pi} \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} = \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} \quad . \tag{45}$$

Also,

$$\hat{\pi}(U \updownarrow) \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} = \sum_k \begin{matrix} |0\rangle\langle 0| \\ |m_k\rangle\langle m_k| \end{matrix} U \begin{matrix} |m_j\rangle \\ |0\rangle \end{matrix} = m_j \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} \quad , \tag{46}$$

and

$$\hat{\pi}(\updownarrow U^\dagger) \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} = \sum_k \begin{matrix} |0\rangle\langle m_k| \\ |m_k\rangle\langle 0| \end{matrix} U^\dagger \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} = m_j^* \begin{matrix} |0\rangle \\ |m_j\rangle \end{matrix} \quad . \tag{47}$$

QED

An immediate consequence of Claim 4 is that

$$\langle m_j 0|U \updownarrow |m_k 0\rangle = \langle m_j 0|\hat{\pi} U \updownarrow |m_k 0\rangle = m_j \delta_j^k \; , \tag{48}$$

for $j, k \in \{0, 1, \ldots, N_S - 1\}$.

Note that since $m_0 = 1$, Eq.(48) implies that

$$|m_0 0\rangle = U \updownarrow |m_0 0\rangle \; . \tag{49}$$

Another consequence of Claim 4 is that $|m_0 0\rangle$ is a stationary state of W. Indeed, one has

$$\begin{aligned} W|m_0 0\rangle &= U(-1)^{\bar{\pi}}U^\dagger(-1)^{\hat{\pi}}|m_0 0\rangle && (50)\\ &= U \updownarrow (1-2\hat{\pi}) \updownarrow U^\dagger(-1)|m_0 0\rangle && (51)\\ &= (1-2m_0 U \updownarrow)(-1)|m_0 0\rangle && (52)\\ &= (1-2)(-1)|m_0 0\rangle && (53)\\ &= |m_0 0\rangle \ . && (54) \end{aligned}$$

Let

$$\mathcal{V}^j_{busy} = span\{|m_j 0\rangle, U \updownarrow |m_j 0\rangle\} \tag{55}$$

for $j \in \{0, 1, \ldots, N_S - 1\}$. (By "span" we mean all linear combinations of these vectors with *complex* coefficients.)

Claim 5 *$W\mathcal{V}^j_{busy} = \mathcal{V}^j_{busy}$ for all $j \in \{0, 1, \ldots, N_S - 1\}$. For $j = 0$, let*

$$|\psi_0\rangle = |m_0 0\rangle \ . \tag{56}$$

$\{|\psi_0\rangle\}$ is an orthonormal basis for $\mathcal{V}^0_{busy}$ and $W|\psi_0\rangle = |\psi_0\rangle$. For $j \neq 0$, let

$$|\psi_{\pm j}\rangle = \frac{\pm i}{\sqrt{2}\sin\varphi_j}(e^{-i\eta_j} U \updownarrow |m_j 0\rangle - e^{\pm i 2\varphi_j}|m_j 0\rangle) \ . \tag{57}$$

$\{|\psi_j\rangle, |\psi_{-j}\rangle\}$ is an orthonormal basis for $\mathcal{V}^j_{busy}$ and $W|\psi_{\pm j}\rangle = e^{\pm i 2\varphi_j}|\psi_{\pm j}\rangle$.

proof:

Using the identities of Claim 4, one finds after some algebra that

$$W|m_j 0\rangle = (-1)|m_j 0\rangle + (2m^*_j) U \updownarrow |m_j 0\rangle \ , \tag{58a}$$

and

$$W(U \updownarrow)|m_j 0\rangle = (-2m_j)|m_j 0\rangle + (-1 + 4|m_j|^2) U \updownarrow |m_j 0\rangle \tag{58b}$$

for all j.

According to Eqs.(58), $\mathcal{V}^j_{busy}$ is invariant under the action of W for each j. By virtue of Eq.(48), $\mathcal{V}^j_{busy}$ is 1-dim for $j = 0$ and 2-dim if $j \neq 0$. We've already proven that $|m_0 0\rangle$ is a stationary state of W.

Now consider a fixed $j \neq 0$. Both $U(-1)^{\bar{\pi}}U^\dagger$ and $(-1)^{\hat{\pi}}$ are reflections, and reflections are a special type of orthogonal matrix, so the product of these 2 orthogonal matrices is also an orthogonal matrix. In fact, it's a rotation about the axis perpendicular to the planar subspace $\mathcal{V}^j_{busy}$. The vectors $|m_j 0\rangle$, and $U \updownarrow |m_j 0\rangle\}$ are

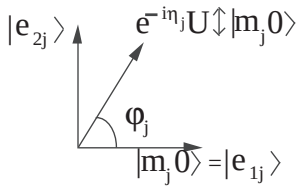


Figure 4: Definition of $|e_{1j}\rangle$ and $|e_{2j}\rangle$.

independent but not orthogonal. However, we can express them in terms of orthogonal vectors (see Fig.4) as follows:

$$|m_j 0\rangle = |e_{1j}\rangle \ , \tag{59a}$$

and

$$e^{-i\eta_j} U \updownarrow |m_j 0\rangle = \cos(\varphi_j)|e_{1j}\rangle + \sin(\varphi_j)|e_{2j}\rangle \ . \tag{59b}$$

In the $|e_{1j}\rangle$, $|e_{2j}\rangle$ basis, we find after substituting $m_j = e^{i\eta_j}\cos(\varphi_j)$ into Eqs.(58) that

$$W = \begin{bmatrix} \cos(2\varphi_j) & \sin(2\varphi_j) \\ -\sin(2\varphi_j) & \cos(2\varphi_j) \end{bmatrix} \ . \tag{60}$$

The eigenvalues of this matrix are $e^{\pm i 2\varphi_j}$, with corresponding eigenvectors:

$$|\psi_{\pm j}\rangle = \frac{1}{\sqrt{2}}(|e_{1j}\rangle \pm |e_{2j}\rangle) \ . \tag{61}$$

These eigenvectors satisfy

$$\langle\psi_{-j}|\psi_j\rangle = 0 \ \ , \ \ \langle\psi_j|\psi_j\rangle = \langle\psi_{-j}|\psi_{-j}\rangle = 1 \ . \tag{62}$$

By expressing $|e_{1j}\rangle$ and $|e_{2j}\rangle$ in Eq.(61) in the original basis, we get Eq.(57).
QED

Define the following vector spaces:

$$\mathcal{V} = span\{|x\rangle \otimes |y\rangle : x, y \in S_{\underline{x}}\} \ , \tag{63}$$

$$\mathcal{V}_A = span\{|x\rangle \otimes |0\rangle : x \in S_{\underline{x}}\} \ , \tag{64}$$

$$\mathcal{V}_B = U \updownarrow \mathcal{V}_A \ , \tag{65}$$

and

$$\mathcal{V}_{busy} = \mathcal{V}_A + \mathcal{V}_B \ . \tag{66}$$

$\mathcal{V}$ can be expressed as a direct sum of $\mathcal{V}_{busy}$ and its orthogonal complement $\mathcal{V}_{busy}^{\perp}$:

$$\mathcal{V} = \mathcal{V}_{busy} \oplus \mathcal{V}_{busy}^{\perp} \ . \tag{67}$$

From Claim 5, it follows that $\mathcal{V}_{busy}$ is a direct sum of the subspaces $\mathcal{V}_{busy}^{j}$:

$$\mathcal{V}_{busy} = \bigoplus_{j=0}^{N_S-1} \mathcal{V}_{busy}^{j} \ . \tag{68}$$

Recall that matrices M_1, M_2 and M_{hyb} are N_S dimensional whereas W is N_S^2 dimensional. Since the size of $S_{\underline{x}}$ is N_S, $dim(\mathcal{V}) = N_S^2$. From Eq.(68) and Claim 5, $dim(\mathcal{V}_{busy}) = 2N_S - 1$. Furthermore, $\{|\psi_j\rangle : j = 0, \pm 1, \pm 2, \ldots, \pm(N_S - 1)\}$ is an orthonormal basis for $\mathcal{V}_{busy}$.

At this point we've explained the action of W on $\mathcal{V}_{busy}$, but we haven't said anything about the action of W on $\mathcal{V}_{busy}^{\perp}$. Next we show that W acts simply as the identity on $\mathcal{V}_{busy}^{\perp}$. (This is what one would expect since the vectors in $\mathcal{V}_{busy}^{\perp}$ are parallel to the axis of the W rotation.) Recall that if S and T are subspaces of a vector space $\mathcal{V}$, then $(S+T)^{\perp} = S^{\perp} \cap T^{\perp}$. Therefore,

$$\mathcal{V}_{busy}^{\perp} = \mathcal{V}_A^{\perp} \cap \mathcal{V}_B^{\perp} \ . \tag{69}$$

From the definitions of $\mathcal{V}_A$ and $\mathcal{V}_B$, it's easy to see that

$$\mathcal{V}_A^{\perp} = span\{|x\rangle \otimes |y\rangle : x \in S_{\underline{x}}, \text{ and } y \in S_{\underline{x}} - \{0\}\} \ , \tag{70}$$

and

$$\mathcal{V}_B^{\perp} = U \updownarrow (\mathcal{V}_A^{\perp}) \ . \tag{71}$$

Claim 6

$$W|\phi\rangle = |\phi\rangle \tag{72}$$

for all $|\phi\rangle \in \mathcal{V}_{busy}^{\perp}$.

proof: Let $|\phi\rangle \in \mathcal{V}_{busy}^{\perp} = \mathcal{V}_A^{\perp} \cap \mathcal{V}_B^{\perp}$. Hence $|\phi\rangle \in \mathcal{V}_A^{\perp}$ and $|\phi\rangle = U \updownarrow |\theta\rangle$ for some $|\theta\rangle \in \mathcal{V}_A^{\perp}$.

$$\begin{aligned} U(-1)^{\check{\pi}}U^{\dagger}(-1)^{\hat{\pi}}|\phi\rangle &= U \updownarrow (-1)^{\hat{\pi}} \updownarrow U^{\dagger}(-1)^{0}|\phi\rangle & (73)\\ &= U \updownarrow (1-2\hat{\pi}) \updownarrow U^{\dagger}U \updownarrow |\theta\rangle & (74)\\ &= U \updownarrow (1-2\hat{\pi})|\theta\rangle & (75)\\ &= |\phi\rangle \ . & (76) \end{aligned}$$

QED

It's interesting to compare the present paper with Ref.[4]. For Ref.[4], $M_1 = M_2 = M$ and $\pi()$ is a standard detailed balance for M instead of a detailed balance for the pair (M_1, M_2). For Ref.[4], $M_{hyb} = M_{sym}$, $m_j = m_j^*$, $U_1 = \check{U}$, $U_2 = \hat{U}$, $U = U_2^\dagger U_1 = \hat{U}^\dagger \check{U}$, $U = \updownarrow U^\dagger \updownarrow$. When $U = \updownarrow U^\dagger \updownarrow$ as in Ref.[4], Eq.(44b) and Eq.(44c) are essentially identical, whereas in the $M_1 \neq M_2$ case, it's less obvious that these two equations are true simultaneously.

5 Quantum Gibbs Sampling Algorithm

In this section, we will describe an algorithm for doing Gibbs sampling on a quantum computer, utilizing the Szegedy operator W that we have so painstakingly discussed in previous sections.

We begin by choosing[2] some $x_0 \in S_{\underline{x}}$ for which $P(\underline{x} = x_0) \neq 0$. Now define

$$|x_0 0\rangle = |\underline{x} = x_0\rangle \otimes |0\rangle^{\otimes N_B} \ . \tag{77}$$

Note that $|x_0 0\rangle \in \mathcal{V}_{busy}$ and

$$\langle \psi_0 | x_0 0\rangle = \langle \sqrt{\pi} | \underline{x} = x_0\rangle = \sqrt{\pi(x_0)} \ . \tag{78}$$

$\sqrt{\pi(x)} = \sqrt{P(x)}$ can be easily evaluated at a single point $x = x_0$. Our quantum Gibbs algorithm consists of performing the original Grover algorithm with beginning state $|x_0 0\rangle$ and target state $|\psi_0\rangle$. Define the following 2 reflection operators

$$R_{beg} = (-1)^{|x_0 0\rangle\langle x_0 0|} \ , \tag{79}$$

and

$$R_{tar} = (-1)^{|\psi_0\rangle\langle\psi_0|} \ . \tag{80}$$

$R_{beg} R_{tar}$ is a rotation by an angle $\mathcal{O}(\sqrt{\pi(x_0)})$ in space $span\{|\psi_0\rangle, |x_0 0\rangle\} \subset \mathcal{V}_{busy}$. Let

$$L = \mathcal{O}(\frac{1}{\sqrt{\pi(x_0)}}) \ . \tag{81}$$

If $\sqrt{\pi(x_0)} = \mathcal{O}(1/\sqrt{N_S})$, then L iterations of $R_{beg} R_{tar}$ will take the beginning state to the target state.[3] To implement this use of Grover's algorithm, we need to compile (with polynomial efficiency) the operator $R_{beg} R_{tar}$. R_{beg} is easy to compile; it's just a single multiply-controlled phase. Next, we will explain how to compile R_{tar}.

[2]Perhaps some symmetry of the physical situation being modeled by the Bayesian network $\underline{x}$ will suggest some $\underline{x}$ value that has non-zero probability. Alternatively, one can proceed as follows. For definiteness, consider a Bayesian net $\underline{x} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)$ with 3 nodes. Suppose $P(x_3, x_2, x_1) = P(x_3|x_2, x_1)P(x_2|x_1)P(x_1)$ and the functions $P_{\underline{x}_3|\underline{x}_2,\underline{x}_1}$ $P_{\underline{x}_2|\underline{x}_1}$ and $P_{\underline{x}_1}$ are known. Choose $y_1 \in S_{\underline{x}_1}$ such that $P_{\underline{x}_1}(y_1) \neq 0$. Then choose $y_2 \in S_{\underline{x}_2}$ such that $P_{\underline{x}_2|\underline{x}_1}(y_2|y_1) \neq 0$. Finally, choose $y_3 \in S_{\underline{x}_3}$ such that $P_{\underline{x}_3|\underline{x}_2,\underline{x}_1}(y_3|y_2, y_1) \neq 0$. Set $x_0 = (y_1, y_2, y_3)$.

[3]We will discuss in a future paper what to do if $\sqrt{\pi(x_0)}$ is much larger than $\mathcal{O}(1/\sqrt{N_S})$.

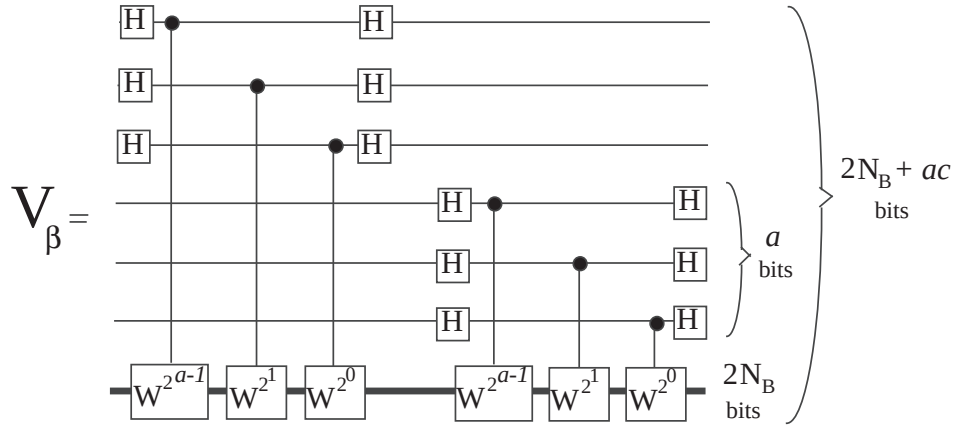


Figure 5: Definition of operator V_β for Szegedy operator W.

Fig.5, which is identical to Fig.18 in Ref.[4], defines an operator V_β in terms of multiple (modified) phase estimation steps. The V_β of Ref.[4] depends on a parameter β (inverse temperature) because the operator M in that paper depends on this parameter. V_β in the present paper does not depend on β (because the Bayesian net $\underline{x}$ doesn't) so we will drop the β subscript from it henceforth, and refer to it simply as V. V does not depend on β but it still depends on the positive integers a and c. (In the language of Ref.[4], a=number of probe bits for each PE (Phase Estimation) step, and c=number of PE steps). Note that operator W is applied $2^a c$ times by V.

Let $|0^{ac}\rangle = |0\rangle^{\otimes(ac)}$, $J = \{0, \pm 1, \pm 2, \ldots, \pm(N_S - 1)\}$, and $J' = J - \{0\}$. According to Lemma 2 of Ref.[3], for any $\epsilon_2 > 0$, if we adjust the integers a and c so that

$$2^a \approx \frac{1}{\Delta} = \mathcal{O}(\frac{1}{\sqrt{\delta}}) \; , \tag{82}$$

and

$$c \approx \log_2(\frac{1}{\sqrt{\epsilon_2}}) \; , \tag{83}$$

(recall $\delta = 1 - |m_1|$ is the distance between the two largest eigenvalue magnitudes of M_1), then V acts on the space $\mathcal{V}_{busy} \otimes |0^{ac}\rangle$ as follows:

$$V = \begin{matrix} |0^{ac}\rangle\langle 0^{ac}| \\ |\psi_0\rangle\langle\psi_0| \end{matrix} + \sum_{j \in J'} \begin{matrix} |\chi_j\rangle\langle 0^{ac}| \\ |\psi_j\rangle\langle\psi_j| \end{matrix} + \mathcal{O}(\sqrt{\epsilon_2}) \; , \tag{84}$$

where the $|\chi_j\rangle$ are states of ac qubits such that $\langle 0^{ac}|\chi_j\rangle = 0$. In Eq.(84) and for the remainder of this section, the top row represents the ac ancilla qubits shown in Fig.5, and the bottom row represents the $2N_B$ qubits on which W operates.

Now define

$$Q = (-1)^{|0^{ac}\rangle\langle 0^{ac}|} = 1 - 2|0^{ac}\rangle\langle 0^{ac}| \; , \tag{85}$$

and

$$\tilde{R}_{tar} = V^\dagger \; \underline{Q} \; V \; . \tag{86}$$

It follows that for any $|\psi\rangle \in \mathcal{V}_{busy}$,

$$\begin{aligned}
\tilde{R}_{tar} \; \begin{matrix} |0^{ac}\rangle \\ |\psi\rangle \end{matrix} &= \left[1 - 2V^\dagger \; \underline{|0^{ac}\rangle\langle 0^{ac}|} \; V \right] \; \begin{matrix} |0^{ac}\rangle \\ |\psi\rangle \end{matrix} && (87) \\
&= \left[1 - 2 \; \begin{matrix} |0^{ac}\rangle\langle 0^{ac}| \\ |\psi_0\rangle\langle\psi_0| \end{matrix} \right] \; \begin{matrix} |0^{ac}\rangle \\ |\psi\rangle \end{matrix} + \mathcal{O}(\sqrt{\epsilon_2}) && (88) \\
&= \begin{matrix} |0^{ac}\rangle \\ |\psi\rangle \end{matrix} + \begin{matrix} |0^{ac}\rangle \\ (-2|\psi_0\rangle\langle\psi_0|)|\psi\rangle \end{matrix} + \mathcal{O}(\sqrt{\epsilon_2}) && (89) \\
&= \begin{matrix} |0^{ac}\rangle \\ R_{tar}|\psi\rangle \end{matrix} + \mathcal{O}(\sqrt{\epsilon_2}) \; . && (90)
\end{aligned}$$

Eq.(90) is the essence of Corollary 2 in Ref.[3]. It means that R_{tar} acting on $\mathcal{V}_{busy}$ can be approximated by $\tilde{R}_{tar}$ acting on $\mathcal{V}_{busy} \otimes |0^{ac}\rangle$. Since we already know how to compile $\tilde{R}_{tar}$, we have accomplished our goal of compiling R_{tar}, at least approximately.

Next, we will try to estimate the error of our quantum Gibbs algorithm. Suppose $\tilde{\pi}()$ is our estimate of $\pi()$. Note that for any $x \in S_{\underline{x}}$,

$$\begin{aligned}
|\pi(x) - \tilde{\pi}(x)| &= |(\sqrt{\pi(x)} - \sqrt{\tilde{\pi}(x)})(\sqrt{\pi(x)} + \sqrt{\tilde{\pi}(x)})| && (91) \\
&\leq 2|\sqrt{\pi(x)} - \sqrt{\tilde{\pi}(x)}| \; . && (92)
\end{aligned}$$

Suppose $\epsilon > 0$ is defined so that

$$\max_x |\sqrt{\pi(x)} - \sqrt{\tilde{\pi}(x)}| \leq \epsilon \; . \tag{93}$$

Then, since we apply the $R_{beg}R_{tar}$ operator a total of L times, and each time we can incur an error of $\sqrt{\epsilon_2}$,

$$\epsilon \approx L\sqrt{\epsilon_2} \; . \tag{94}$$

If we define one step as one W application, then the total number of steps for the whole algorithm is $\mathcal{O}(L2^a c) = \mathcal{O}(\frac{L}{\sqrt{\delta}} \log_2(\frac{L}{\epsilon}))$. Thus, our algorithm will yield a sample of the classical Bayesian net $\underline{x}$ with precision $\mathcal{O}(\epsilon)$, in $\mathcal{O}(\frac{L}{\sqrt{\delta}} \log_2(\frac{L}{\epsilon}))$ steps. Achieving the same precision with a classical Gibbs sampling algorithm would require $\mathcal{O}(\frac{1}{\delta})$ steps.

The Szegedy operator W of this paper can also be used to do quantum simulated annealing and Metropolis-Hastings if the marginals $P(x_i^{t+1}|x_{\{i\}^c}^t)$ can be calculated for each i from the transition matrix $P(x^{t+1}|x^t)$. (In the case of simulated annealing, $P(x^{t+1}|x^t)$ is different for each β_i of the annealing schedule).

References

[1] Mario Szegedy, "Spectra of Quantized Walks and a $\sqrt{\delta\epsilon}$ rule", arXiv:quant-ph/0401053

[2] R. Somma, S. Boixo, H. Barnum, "Quantum Simulated Annealing", arXiv:0712.1008

[3] Pawel Wocjan, Anura Abeyesinghe, "Speed-up via Quantum Sampling", arXiv:0804.4259

[4] R.R. Tucci, "Code Generator for Quantum Simulated Annealing", arXiv:0908.1633

[5] R.R. Tucci, "Use of a Quantum Computer to do Importance and Metropolis-Hastings Sampling of a Classical Bayesian Network", arXiv:0811.1792